Comprehensive Resequence Analysis of a 97 kb Region of Chromosome 10q11.2 Containing the MSMB Gene Associated with Prostate Cancer

Hum Genet (2009) 126:743–750
DOI 10.1007/s00439-009-0723-9

ORIGINAL INVESTIGATION

Comprehensive resequence analysis of a 97 kb region of chromosome 10q11.2 containing the *MSMB* gene associated with prostate cancer

Meredith Yeager · Zuoming Deng · Joseph Boland · Casey Matthews · Jennifer Bacior · Victor Lonsberry · Amy Hutchinson · Laura A. Burdett · Liqun Qi · Kevin B. Jacobs · Jesus Gonzalez-Bosquet · Sonja I. Berndt · Richard B. Hayes · Robert N. Hoover · Gilles Thomas · David J. Hunter · Michael Dean · Stephen J. Chanock

Received: 30 June 2009 / Accepted: 17 July 2009 / Published online: 31 July 2009

Abstract Genome-wide association studies of prostate cancer have identified single nucleotide polymorphism (SNP) markers in a region of chromosome 10q11.2, harboring the microseminoprotein-β (*MSMB*) gene. Both the gene product of *MSMB*, the prostate secretory protein 94 (PSP94) and its binding protein (PSPBP), have been previously investigated as serum biomarkers for prostate cancer progression. Recent functional work has shown that different alleles of the significantly associated SNP in the promoter of *MSMB* found to be associated with prostate cancer risk, rs10993994, can influence its expression in tumors and in vitro studies. Since it is plausible that additional variants in this region contribute to the risk of prostate cancer, we have used next-generation sequencing technology to resequence a ~97-kb region that includes the area surrounding *MSMB* (chr10: 51,168,025–51,265,101) in 36 prostate cancer cases, 26 controls of European origin, and 8 unrelated CEPH individuals in order to identify additional variants to investigate in functional studies. We identified 241 novel polymorphisms within this region, including 142 in the 51-kb block of linkage disequilibrium (LD) that contains rs10993994 and the proximal promoter of *MSMB*. No sites were observed to be polymorphic within the exons of *MSMB*.

Introduction

Genome-wide association studies (GWAS) have been instrumental in identifying novel regions of the genome that are associated with human diseases and traits (Donnelly 2008; Manolio et al. 2008). Though prostate cancer has a

Electronic supplementary material The online version of this article (doi:10.1007/s00439-009-0723-9) contains supplementary material, which is available to authorized users.

M. Yeager · Z. Deng · J. Boland · C. Matthews · J. Bacior · V. Lonsberry · A. Hutchinson · L. A. Burdett · L. Qi
Core Genotyping Facility, Advanced Technology Program, SAIC-Frederick, Inc., NCI-Frederick, Frederick, MD 21702, USA

M. Yeager · Z. Deng · J. Boland · C. Matthews · J. Bacior · V. Lonsberry · A. Hutchinson · L. A. Burdett · L. Qi · K. B. Jacobs · J. Gonzalez-Bosquet · S. I. Berndt · R. B. Hayes · R. N. Hoover · G. Thomas · S. J. Chanock
Division of Cancer Epidemiology and Genetics, NCI, NIH, Bethesda, MD 20892, USA

K. B. Jacobs
Bioinformed Consulting Services, Gaithersburg, MD 20877, USA

R. B. Hayes
Division of Epidemiology, New York University Langone Medical Center, New York, NY 10016, USA

D. J. Hunter
Program in Molecular and Genetic Epidemiology, Department of Epidemiology, Harvard School of Public Health, Boston, MA 02115, USA

M. Dean
Laboratory of Experimental Immunology, Cancer and Inflammation Program, Center for Cancer Research, NCI-Frederick, National Cancer Institute, Frederick, MD 21702, USA

M. Yeager (✉)
Advanced Technology Center, NCI, 8717 Grovemont Circle, Gaithersburg, MD 20877, USA
e-mail: yeagerm@mail.nih.gov

high incidence rate (~170 per 100,000 men in the United States) (Crawford 2003), until recently, the only well-established risk factors for prostate cancer were male sex, African ancestry, age, and family history (Johns and Houlston 2003). Genetic factors have been postulated to contribute to disease, and twin studies have shown high heritability (Lichtenstein et al. 2000). However, candidate gene association studies and early linkage analyses yielded inconclusive results. Recently, GWAS have identified at least 20 common genetic variants associated with prostate cancer risk (Amundadottir et al. 2006; Eeles et al. 2008; Freedman et al. 2006; Gudmundsson et al. 2007a, b; Haiman et al. 2007; Thomas et al. 2008; Yeager et al. 2007). These findings have forged new avenues for prostate cancer researchers to investigate regions that contribute to prostate cancer risk through yet to be defined mechanisms.

Two independent prostate cancer GWAS identified a single-nucleotide polymorphism (SNP; rs10993994 C>T) on chromosome 10q11.2, which maps to the proximal promoter of the microseminoprotein-β gene (*MSMB*) (Eeles et al. 2008; Thomas et al. 2008). The protein encoded by *MSMB*, prostate secretory protein 94 (PSP94), and its binding protein (PSPBP), have been evaluated as markers for early detection (Nam et al. 2006) and prognosis of prostate cancer (Bjartell et al. 2007; Reeves et al. 2006). The T allele of rs10993994 is significantly associated with the risk of prostate cancer ($P = 9.7 \times 10^{-19}$; OR = 1.20) (Lou et al. 2009), and follow-up work has shown evidence that alleles of rs10993994 differentially influence the expression of *MSMB* (Chang et al. 2009; Lou et al. 2009). While the C allele was shown to be required for *MSMB* expression, the T allele was associated with lower expression levels of *MSMB* in vitro (Lou et al. 2009). Interestingly, these results are consistent with the observation that *MSMB* expression levels progressively decrease during the development of prostate cancer from early to late stages (LaTulippe et al. 2002; Stanbrough et al. 2006; Vanaja et al. 2003). Together, these data provide a plausible mechanism for the contribution of the T allele to prostate cancer risk.

While there is strong evidence that variation at rs10993994 influences both *MSMB* expression and prostate cancer risk, it is possible that other genetic variants within this region of chromosome 10q14 could contribute to risk of developing prostate cancer, perhaps altering *MSMB* expression or through other mechanisms (Lou et al. 2009). Without a comprehensive assessment of common genetic variation across this region of 10q11.2, it is not possible to define the set of high priority common variants suitable for follow-up studies. In order to comprehensively catalog common genetic variation across *MSMB* and the neighboring region defined by the structure of the linkage disequilibrium, we used Roche-454 next-generation sequencing technology (Rothberg and Leamon 2008; Yeager et al. 2008) to resequence, in 70 unrelated individuals, a 97-kb region of chromosome 10 (chr10: 51,168,025–51,265,101) that includes a large block (51 kb) of linkage disequilibrium (LD) in which rs10993994 resides, *MSMB*, and a neighboring gene, nuclear receptor coactivator 4 (*NCO4*), that is also a possible candidate gene for prostate cancer risk.

Materials and methods

Samples

For sequence analysis, 38 prostate cancer cases and 27 controls were selected from the National Cancer Institute's (NCI) prostate, lung, colorectal, and ovarian (PLCO) cancer screening trial (Gohagan et al. 2000); all (97%) but two were included in the initial prostate cancer GWAS (Yeager et al. 2007) that was previously conducted as a part of the Cancer GEnetics Markers of Susceptibility (CGEMS; see http://cgems.cancer.gov/) initiative. Additionally, 11 individuals from CEPH pedigrees were used in the analysis, including two sets of relatives. Seven of these individuals were CEU samples from The International HapMap Project (http://hapmap.org).

Region selection

Using CEU data obtained from the International HapMap Project (http://www.hapmap.org; The International HapMap Consortium 2005), based on the pattern of linkage disequilibrium (LD) flanking rs10993994 using Haploview (Barrett et al. 2005), we targeted 97 kb (chr10: 51,167,932–51,265,146), which included a 51-kb block of LD centromeric of *MSMB* with *r*s10993994 at its most telomeric position but none of the coding exons. We extended our target sequencing by an additional 46 kb to include *MSMB* and the telomeric neighboring gene, nuclear receptor coactivator 4 (*NCOA4*).

PCR primer design

Twenty sets of long-range PCR primers were designed to cover the 97-kb targeted region. Approximate amplicon size ranged from 3,544 to 5,542 bp; primer sets were designed to overlap, and averaged 292 bp overlap with the adjacent primer set. Primer design was as follows: primers had been initially designed using Primer3 (http://frodo.wi.mit.edu) (Rozen and Skaletsky 2000), were then quality-checked in silico for uniqueness, potential sequence paralogy and DNA repeat sequences using the BLAT feature of the UCSC genome browser (http://genome.ucsc.edu/cgi-bin/hgBlat). Next, NetPrimer (http://www.premierbiosoft.com/netprimer/index.html) was used to check for primer secondary structures

and efficiencies. Primers were ordered from Integrated DNA Technologies (Coralville, IA, USA; http://www.idtdan.com). All primers, coordinates, and PCR conditions are supplied in Supplementary Table 1.

Sequencing

Post long-range PCR, all protocols were followed in accordance to standard kits used for the 454 GS FLX system (http://www.454.com/products-solutions/product-list.asp).

Polymorphism detection

An automated computational pipeline was developed to process sequence reads generated by 454 FLX genome sequencers. Whenever applicable, sequence reads from the same sample were pooled based on barcodes provided by Roche/454. QC was performed with vendor-supplied software, and sequence reads that passed QC were aligned to the target genomic region (chr10: 51,167,932–51,265,146) by MOSAIK software (http://bioinformatics.bc.edu/marthlab/Mosaik). The resulting assembly was analyzed in a column-by-column approach and potential polymorphic sites and most likely genotypes were called based on a set of heuristic rules. For each nucleotide position, the minimal sequence coverage depth was set to 20 reads. In addition, the ratio (r) of forward and reverse reads was determined. To avoid directional bias, an optimal range of r was set between 10 and 90%. Homozygous genotype calls were made when the most frequent allele was present in at least 85% of the reads. Heterozygous genotype calls were made when the two most frequent alleles were represented in 30–70% of reads. No genotype calls were made if the above criteria were not met. Manual inspections aided by the NextGENe software (http://www.softgenetics.com) and Consed (http://bozeman.mbt.washington.edu/consed/consed.html) were performed to QA the results and resolve ambiguous cases.

Descriptive statistics

Completion, concordance, minor allele frequency (MAF) estimations, deviations from fitness for Hardy–Weinberg proportion (HWP), pair-wise LD, and tag SNP selection were computed using the GLU software package (http://code.google.com/p/glu-genetics/).

Results

Coverage and depth

Excellent coverage and depth was obtained across the entire sequenced region. Figure 1 shows the coverage and depth averaged across all samples. The average depth was 104X, with a minimum of 24X; in two small regions (chr10: 51,177,997–51,183,539 and 51,241,521–51,246,917) that correspond to two amplicons we consistently observe an average coverage depth of <50X, though in both regions coverage did not fall below 25X.

We consistently observed over-representation of sequence reads at the ends of amplicons, regardless of the genomic regions or individual samples sequenced. The coverage variability was as high as three or more times the average, an artifact that is most likely introduced during the shotgun sequencing library preparation of PCR amplicons (Harismendy and Frazer 2009). Another contributing factor to the "spikes" in Fig. 1 is that adjacent amplicons are usually overlapped at the ends, which increases the chance of sequence oversampling.

Polymorphism detection and genotype and sample quality control

The heuristic rules we implemented ("Materials and methods") took into consideration uneven sequence coverage. Therefore, coverage spikes at the amplicon ends (Fig. 1) have minimal impact on genotype accuracy. Genotypes were initially called for 685 SNPs and insertion/deletion polymorphisms (indels) in 76 individuals, including sites that had been previously reported in the NCBI dbSNP database (http://www.ncbi.nlm.nih.gov/projects/SNP/). Iterations of data cleaning excluded two related CEPH individuals and 41 loci with significant genotype issues (e.g., failure of Hardy–Weinberg proportions; $P < 0.001$) and/or with very low (<50%) genotype calling (completion) rates; 11 of these 41 polymorphisms fell within the chr10: 51,241,521–51,246,917 region of lower average coverage as discussed above. Within the other region of lower-than-average coverage, genotypes were successfully obtained for 18 loci, and the observed heterozygosity within this interval did not differ than the average across other regions. Of the remaining 637 loci, 190 sites in dbSNP were observed to be monomorphic within our samples and were excluded for subsequent analyses.

Utilizing the CGEMS prostate follow-up #2 iSelect (manuscripts in progress), we examined the 27 SNPs that overlapped the sequence data in 41 PLCO samples. In three samples, genotype concordance rates were <75%, all of which had low completion rates; all other samples were 100% concordant. After removing these three samples, genotype concordance per locus was 100% (Supplementary Table 2). Data from the International HapMap Project (http://www.hapmap.org) were used for 70 SNPs genotyped within this region over the seven HapMap CEU individuals. One sample displayed <95% concordance and was removed from further analyses. Concordance for two

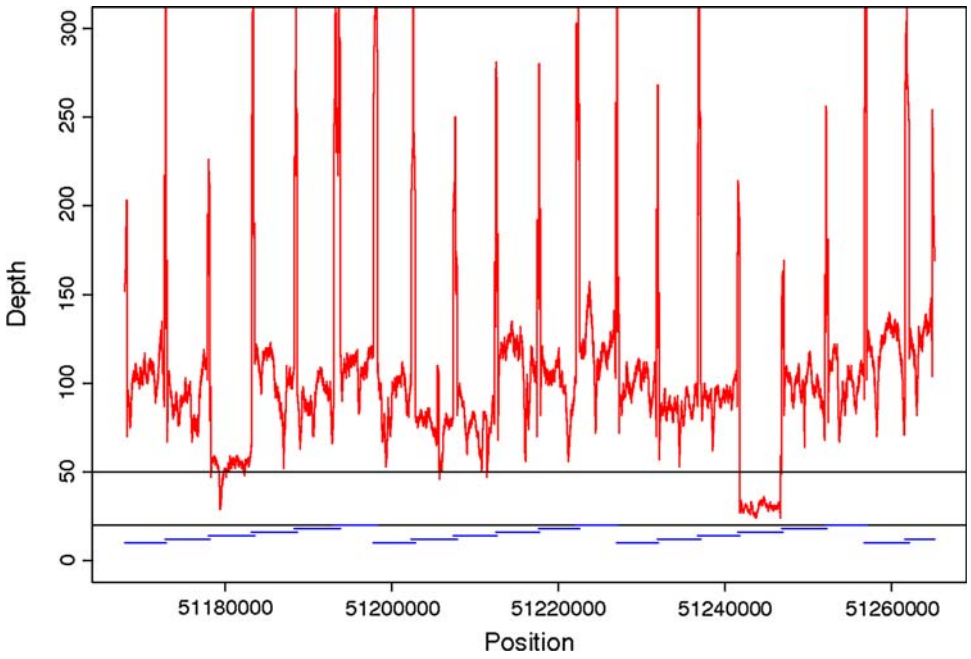

Fig. 1 Coverage and depth averaged over all samples for the 97-kb region of chromosome 10q11.2. The region consists of chr10: 51,167,932–51,265,146), which includes a 51-kb block of LD centromeric of *MSMB* with *r*s10993994 at its most telomeric position

SNPs was problematic in the remaining six samples with all individuals displaying discordant genotypes; these two SNPs were excluded from further analyses. Three other SNPs displayed one discordant genotype each, and the remaining 65 SNPs displayed 100% concordance between HapMap genotypes and genotypes derived by sequencing.

Final dataset statistics

The final genotype dataset contained 70 individuals (36 and 26 PLCO prostate cancer cases and controls, respectively, and 8 unrelated CEPH individuals) and 440 polymorphic loci. The average call rate for genotypes was 92% (range 50–100%; median 97%; see Supplementary Table 2). Minor allele frequency (MAF) estimates were computed for each SNP and indel and overall averaged 15% (range 0.7–50%; see Supplementary Table 2); MAF by chromosomal position is shown in Fig. 2. Overall, 181 sites (169 SNPs and 12 indels) were observed to have a MAF $\geq$ 5% (Table 1). A total of 115 SNPs and 44 indels were observed in only one chromosome each. Since our indel-calling algorithm is currently still being refined, these low-frequency variants should be treated as preliminary; we include them here for completeness of probable polymorphic sites.

Table 1 shows the classification of SNPs and indels with respect to previous dbSNP, HapMap, and Illumina HumanHap610 inclusion. Of the 440 polymorphic loci, 199 had previously been reported in the dbSNP database, only 84 of which contained available frequency information in dbSNP at the time of analysis (see Table 1). Within this region, 68 SNPs were genotyped as a part of the HapMap

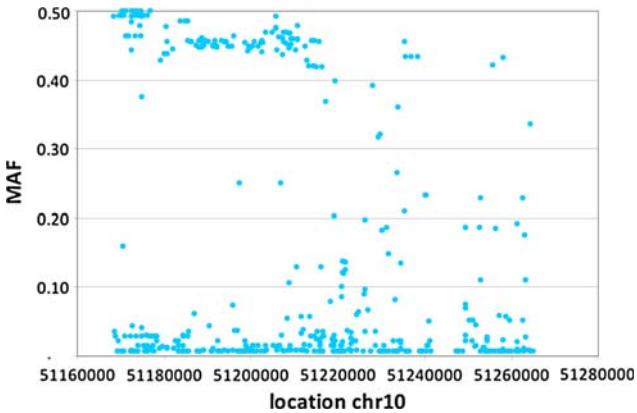

Fig. 2 Minor allele frequencies by position for all polymorphic variants across the 97 kb of chromosome 10q11.2. All polymorphic sites are included across chr10: 51,167,932–51,265,146

Project (http://www.hapmap.org), and 16 SNPs are included on the Illumina HumanHap610 fixed-content GWAS chip. In this study, we observed 241 novel SNPs and indels, of which 98 SNPs and 44 indels were observed as singletons and many others relatively rarely (<5% MAF).

No novel polymorphic sites were observed within the exons of the *MSMB* gene, nor were any of the exonic SNPs that had been previously reported ($N = 24$ in dbSNP) observed to be polymorphic within our samples. For *NCOA4*, no sites were observed to be polymorphic within the first six exons. However, within exons 7, 8, and 10 there were 2, 3, and 4 polymorphic sites observed, respectively. Supplementary Table 3 includes the subset of variants observed within the exons of the *NCOA4* gene. Of the observed polymorphic sites, two were synonymous

Table 1 Distribution of polymorphic SNPs and indels with regards to Illumina HumanHap610, dbSNP and HapMap inclusion

Content[a]	SNPs		Indels		% bin coverage	Untagged bins[c]
	All	>5%[b]	All	>5%[b]		
dbSNP 127[a]	191	154	8	5	83.3	6
No prior frequency information	108	84	7	4	55.6	16
Prior frequency information	83	70	1	1	55.6	16
HapMap I + II[a]	68	58	0	0	50	18
Illumina HumanHap610[a]	16	16	0	0	27.8	26
Novel	157	15	84	7	33.3	24
All variants	348	169	93	12	100	0

169 dbSNP SNPs were found to be monomorphic within our samples

[a] Illumina HumanHap610, HapMap I + II, and dbSNP SNPs overlap since all SNPs are included in dbSNP

[b] Minor allele frequency, MAF

[c] Number of additional tags required to monitor the remaining variation

changes, tree non-synonymous, and four fell within the 3′UTR of the gene; all but two of the SNPs were observed in only one chromosome apiece (rs41306524, D223D; rs11548236, 3′UTR).

Linkage disequilibrium and tag SNP selection

An extremely high degree of LD was observed among polymorphic sites within the first 51 kb of the 97 kb sequenced region (Fig. 3; values shown are r^2). This region contains the previously associated functional SNP, rs10993994, as well as the proximal promoter of *MSMB*. Outside of this large block, significant LD was not observed within the region containing *MSMB* and *NCOA4*.

Tagging analysis of the 97-kb region at an r^2 of 0.8 with rs10993994 as an obligate-include gave a total of 36 tags necessary to cover the 181 loci with a MAF > 5% in individuals of European background. The bin containing rs10993994 contains 20 other SNPs but no indels. Supplementary Table 4 shows the pairwise r^2 of these SNPs ranges from 0.8 to 0.85. It is notable that no other variant was observed to be completely correlated with rs10993994. Tagging at more aggressive r^2 thresholds of 0.9 and 1.0 increased the number of tags to 39 and 116, respectively. Supplementary Table 5 lists the bin and tag SNP information using three thresholds: $r^2 \geq 0.8$, $r^2 \geq 0.9$, and $r^2 = 1.0$.

In order to determine the overall contribution of the newly described variants from this study with respect to coverage of the 97 kb region, we tagged ($r^2 \geq 0.8$, minimum MAF = 5%) based on prior inclusion of SNPs and indels in dbSNP, both (1) with and (2) without reported frequency information, (3) inclusion in HapMap, (4) inclusion on the Illumina HumanHap610 fixed-panel GWAS chip, and (5) novel variants discovered as a part of this effort. In total, 154 SNPs and 5 indels reported in dbSNP (http://www.ncbi.nlm.nih.gov/SNP) provide 83% bin coverage (Table 1) of the region. However, since the majority of these SNPs ($N = 108$) contained no frequency information within the database, this study provides further characterization for these predicted variants. A subset ($N = 58$) of dbSNP SNPs were characterized as a part of the International HapMap Project (http://www.hapmap.org), and only provide 50% bin coverage of the 97-kb region. An even smaller subset of SNPs ($N = 16$) are included on the widely used Illumina HumanHap610 fixed

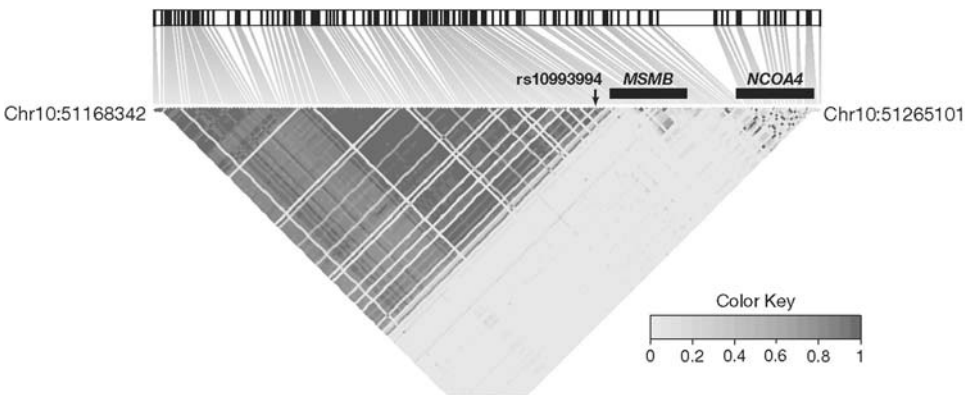

Fig. 3 Linkage disequilibrium (LD) of polymorphisms ≥5% MAF as measured by r^2 across a 97-kb region of chromosome 10q11.2. Relative locations of rs10993994, *MSMB*, and *NCOA4* are shown, though not to scale. Coordinates based on NCBI genome build 36.3

panel GWAS chip. These 16 SNPs only provide 28% bin coverage. Considering only the 22 common (>5% MAF) SNPs discovered by our resequencing project, these novel variants provide 33% bin coverage. All SNPs and indels taken together, regardless of source, provide 100% coverage of this region, and will be extremely useful in studies focused on this region of the genome.

Discussion

In this study, we have characterized the set of common SNP and indel variants across a 97-kb region of chromosome 10q11.2, and in the process have doubled the number of variants available for follow-up in individuals of European background. It is notable that we did not discover additional SNPs or indels in perfect LD with rs10993994, the promoter SNP most strongly associated with prostate cancer risk in two independent GWAS, which has also been shown to display a molecular phenotype consistent with decreased expression of *MSMB*. Consequently, our findings further point toward rs10993994 as the probable variant that at minimum partially accounts for the association with prostate cancer. Our analysis also provides a comprehensive portrait of common variation across this region. Tagging of the region with an r^2 threshold of 0.8 for variants with a MAF > 5%, 36 variants are required, but tagging at more aggressive r^2 thresholds of 0.9 and 1.0 increased the number of tags to 39 and 116, respectively.

There is increasing evidence that the microseminoprotein-β (*MSMB*) gene on chromosome 10q11.2 plays a role in the development of prostate cancer. GWAS studies have pointed to this region and follow-up studies have shown that altered gene expression of *MSMB*, as a consequence of the most significantly associated variants in the promoter, correlated with prostate cancer risk. Levels of *MSMB* reportedly are decreased in patients with poorer outcome and higher cancer recurrence (Nam et al. 2006; Reeves et al. 2006). Furthermore, the protein encoded by *MSMB* has been shown to suppress the growth and induce apoptosis of prostate cancer cells (Lokeshwar et al. 1993; Mundle and Sheth 1993) Though strong evidence suggests that alleles of rs10993994 are functionally important (Chang et al. 2009; Lou et al. 2009) and account for a portion of prostate cancer susceptibility (Chang et al. 2009; Eeles et al. 2008; Lou et al. 2009; Thomas et al. 2008), it is also plausible that other variants in this region of chromosome 10q11.2, perhaps in other genes such as *NCO4*, could further contribute to risk (Lou et al. 2009).

Using 454 next-generation sequencing technology (Rothberg and Leamon 2008), we have extensively characterized common genetic variation across this region of chromosome 10q11.2, which includes the *MSMB* gene and a large upstream block of LD that contains its proximal promoter and the SNP for which functional alleles have been described, rs10993994 (Chang et al. 2009; Lou et al. 2009). Though 24 exonic variants are reported in the dbSNP database, no polymorphic variants were confirmed within the exons of the *MSMB* gene. Within the ~51-kb block of LD that contains rs10993994 and the proximal promoter of *MSMB*, we characterized 282 genetic variants, including 94 and 49 novel SNPs and indels, respectively.

As a part of this project, we also included the sequencing of a neighboring gene, nuclear receptor coactivator 4 (*NCOA4*, also known as ARA70). *MSMB* and *NCOA4* are physically located ~10 kb apart, and *NCOA4* may also be of interest in prostate cancer since it is a coregulator of androgen receptor (AR) (Yeh and Chang 1996). The ARA70 protein has at least two known isoforms (α and β), both of which may have effects on prostate cancer growth though through potentially different mechanisms. ARA70α inhibits androgen-dependent cell proliferation, and ARA70β promotes proliferation and enhances prostate cancer invasion (Peng et al. 2008). In addition to observing an association with rs10993994, Chang et al. (2009) observed a potential second region of association (represented by rs10761581) in a block of LD across *NCOA4* that was separated from rs10993994 by a recombination hotspot (Chang et al. 2009), though the association was not consistent across all replication studies. Common genetic variation in this region should be further explored for independent evidence of association with prostate cancer; a precedent has already been set for multiple independent associations in close proximity from prostate cancer GWAS, including 8q24 (Gudmundsson et al. 2007a; Haiman et al. 2007; Yeager et al. 2007) and *HNF1B* (Sun et al. 2008).

The relationship between the two contiguous genes on chromosome 10q11.2, *MSMB* and *NCO4* could be complex. In the National Center for Biotechnology Information (NCBI) public database AceView (http://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/) (Thierry-Mieg and Thierry-Mieg 2006)), also displayed within the UCSC genome browser (http://genome.ucsc.edu/) (Kent et al. 2002), *MSMB* and *NCOA4* form a complex locus in which mRNA transcripts utilize exons from both genes. For example, expressed sequence tag (EST) DB215804.1 (Kimura et al. 2006) is comprised of exons 1 and 2 of *MSMB* and exons 2–4 of *NCOA4*. Therefore, the *MSMB* promoter could affect the regulation of hybrid MSMB-NCOA4 proteins. It will be important to further explore the role of genetic variation at *NCOA4* and prostate cancer risk.

We have characterized common genetic variation across ~97 kb of chromosome 10 (chr10: 51,168,025–51,265,101) that is of immediate interest in prostate cancer research, since it is comprised of (1) a region of linkage disequilibrium (LD)

in which a known functional polymorphism (rs10993994) resides, (2) a strong candidate gene for prostate cancer risk (*MSMB*), and (3) an additional gene (*NCOA4*) that is of interest in prostate cancer and parts of which that may, in some cases, be transcribed with exons from *MSMB*. Common and uncommon variants identified in this comprehensive resequence analysis of a region of chromosome 10q11.2 have established a foundation for choosing tag SNPs in follow-up studies to further map the region in an effort to nominate the most likely candidate variants for follow-up functional laboratory work to investigate the molecular basis of prostate cancer risk in this region.

Acknowledgments This project has been funded in whole or in part with federal funds from the National Cancer Institute, National Institutes of Health, under Contract No. HHSN261200800001E. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

References

Amundadottir LT, Sulem P, Gudmundsson J, Helgason A, Baker A, Agnarsson BA, Sigurdsson A, Benediktsdottir KR, Cazier JB, Sainz J, Jakobsdottir M, Kostic J, Magnusdottir DN, Ghosh S, Agnarsson K, Birgisdottir B, Le Roux L, Olafsdottir A, Blondal T, Andresdottir M, Gretarsdottir OS, Bergthorsson JT, Gudbjartsson D, Gylfason A, Thorleifsson G, Manolescu A, Kristjansson K, Geirsson G, Isaksson H, Douglas J, Johansson JE, Balter K, Wiklund F, Montie JE, Yu X, Suarez BK, Ober C, Cooney KA, Gronberg H, Catalona WJ, Einarsson GV, Barkardottir RB, Gulcher JR, Kong A, Thorsteinsdottir U, Stefansson K (2006) A common variant associated with prostate cancer in European and African populations. Nat Genet 38:652–658

Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21:263–265

Bjartell AS, Al-Ahmadie H, Serio AM, Eastham JA, Eggener SE, Fine SW, Udby L, Gerald WL, Vickers AJ, Lilja H, Reuter VE, Scardino PT (2007) Association of cysteine-rich secretory protein 3 and beta-microseminoprotein with outcome after radical prostatectomy. Clin Cancer Res 13:4130–4138

Chang BL, Cramer SD, Wiklund F, Isaacs SD, Stevens VL, Sun J, Smith S, Pruett K, Romero LM, Wiley KE, Kim ST, Zhu Y, Zhang Z, Hsu FC, Turner AR, Adolfsson J, Liu W, Kim JW, Duggan D, Carpten J, Zheng SL, Rodriguez C, Isaacs WB, Gronberg H, Xu J (2009) Fine mapping association study and functional analysis implicate a SNP in MSMB at 10q11 as a causal variant for prostate cancer risk. Hum Mol Genet 18(7):1368–1375

Crawford ED (2003) Epidemiology of prostate cancer. Urology 62:3–12

Donnelly P (2008) Progress and challenges in genome-wide association studies in humans. Nature 456:728–731

Eeles RA, Kote-Jarai Z, Giles GG, Olama AA, Guy M, Jugurnauth SK, Mulholland S, Leongamornlert DA, Edwards SM, Morrison J, Field HI, Southey MC, Severi G, Donovan JL, Hamdy FC, Dearnaley DP, Muir KR, Smith C, Bagnato M, Ardern-Jones AT, Hall AL, O'Brien LT, Gehr-Swain BN, Wilkinson RA, Cox A, Lewis S, Brown PM, Jhavar SG, Tymrakiewicz M, Lophatananon A, Bryant SL, Horwich A, Huddart RA, Khoo VS, Parker CC, Woodhouse CJ, Thompson A, Christmas T, Ogden C, Fisher C, Jamieson C, Cooper CS, English DR, Hopper JL, Neal DE, Easton DF (2008) Multiple newly identified loci associated with prostate cancer susceptibility. Nat Genet 40:316–321

Freedman ML, Haiman CA, Patterson N, McDonald GJ, Tandon A, Waliszewska A, Penney K, Steen RG, Ardlie K, John EM, Oakley-Girvan I, Whittemore AS, Cooney KA, Ingles SA, Altshuler D, Henderson BE, Reich D (2006) Admixture mapping identifies 8q24 as a prostate cancer risk locus in African-American men. Proc Natl Acad Sci USA 103:14068–14073

Gohagan JK, Prorok PC, Hayes RB, Kramer BS (2000) The prostate, lung, colorectal and ovarian (PLCO) cancer screening trial of the National Cancer Institute: history, organization, and status. Control Clin Trials 21:251S–272S

Gudmundsson J, Sulem P, Manolescu A, Amundadottir LT, Gudbjartsson D, Helgason A, Rafnar T, Bergthorsson JT, Agnarsson BA, Baker A, Sigurdsson A, Benediktsdottir KR, Jakobsdottir M, Xu J, Blondal T, Kostic J, Sun J, Ghosh S, Stacey SN, Mouy M, Saemundsdottir J, Backman VM, Kristjansson K, Tres A, Partin AW, Albers-Akkers MT, Godino-Ivan Marcos J, Walsh PC, Swinkels DW, Navarrete S, Isaacs SD, Aben KK, Graif T, Cashy J, Ruiz-Echarri M, Wiley KE, Suarez BK, Witjes JA, Frigge M, Ober C, Jonsson E, Einarsson GV, Mayordomo JI, Kiemeney LA, Isaacs WB, Catalona WJ, Barkardottir RB, Gulcher JR, Thorsteinsdottir U, Kong A, Stefansson K (2007a) Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. Nat Genet 39:631–637

Gudmundsson J, Sulem P, Steinthorsdottir V, Bergthorsson JT, Thorleifsson G, Manolescu A, Rafnar T, Gudbjartsson D, Agnarsson BA, Baker A, Sigurdsson A, Benediktsdottir KR, Jakobsdottir M, Blondal T, Stacey SN, Helgason A, Gunnarsdottir S, Olafsdottir A, Kristinsson KT, Birgisdottir B, Ghosh S, Thorlacius S, Magnusdottir D, Stefansdottir G, Kristjansson K, Bagger Y, Wilensky RL, Reilly MP, Morris AD, Kimber CH, Adeyemo A, Chen Y, Zhou J, So WY, Tong PC, Ng MC, Hansen T, Andersen G, Borch-Johnsen K, Jorgensen T, Tres A, Fuertes F, Ruiz-Echarri M, Asin L, Saez B, van Boven E, Klaver S, Swinkels DW, Aben KK, Graif T, Cashy J, Suarez BK, van Vierssen Trip O, Frigge ML, Ober C, Hofker MH, Wijmenga C, Christiansen C, Rader DJ, Palmer CN, Rotimi C, Chan JC, Pedersen O, Sigurdsson G, Benediktsson R, Jonsson E, Einarsson GV, Mayordomo JI, Catalona WJ, Kiemeney LA, Barkardottir RB, Gulcher JR, Thorsteinsdottir U, Kong A, Stefansson K (2007b) Two variants on chromosome 17 confer prostate cancer risk, and the one in TCF2 protects against type 2 diabetes. Nat Genet 39:977–983

Haiman CA, Patterson N, Freedman ML, Myers SR, Pike MC, Waliszewska A, Neubauer J, Tandon A, Schirmer C, McDonald GJ, Greenway SC, Stram DO, Le Marchand L, Kolonel LN, Frasco M, Wong D, Pooler LC, Ardlie K, Oakley-Girvan I, Whittemore AS, Cooney KA, John EM, Ingles SA, Altshuler D, Henderson BE, Reich D (2007) Multiple regions within 8q24 independently affect risk for prostate cancer. Nat Genet 39:638–644

Harismendy O, Frazer K (2009) Method for improving sequence coverage uniformity of targeted genomic intervals amplified by LR-PCR using Illumina GA sequencing-by-synthesis technology. Biotechniques 46:229–231

Johns LE, Houlston RS (2003) A systematic review and meta-analysis of familial prostate cancer risk. BJU Int 91:789–794

Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D (2002) The human genome browser at UCSC. Genome Res 12:996–1006

Kimura K, Wakamatsu A, Suzuki Y, Ota T, Nishikawa T, Yamashita R, Yamamoto J, Sekine M, Tsuritani K, Wakaguri H, Ishii S, Sugiyama T, Saito K, Isono Y, Irie R, Kushida N, Yoneyama T, Otsuka R, Kanda K, Yokoi T, Kondo H, Wagatsuma M, Murakawa K, Ishida S, Ishibashi T, Takahashi-Fujii A, Tanase T, Nagai K, Kikuchi H, Nakai K, Isogai T, Sugano S (2006) Diversification of transcriptional modulation: large-scale identification and characterization of putative alternative promoters of human genes. Genome Res 16:55–65

LaTulippe E, Satagopan J, Smith A, Scher H, Scardino P, Reuter V, Gerald WL (2002) Comprehensive gene expression analysis of prostate cancer reveals distinct transcriptional programs associated with metastatic disease. Cancer Res 62:4499–4506

Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, Pukkala E, Skytthe A, Hemminki K (2000) Environmental and heritable factors in the causation of cancer—analyses of cohorts of twins from Sweden, Denmark, and Finland. N Engl J Med 343:78–85

Lokeshwar BL, Hurkadli KS, Sheth AR, Block NL (1993) Human prostatic inhibin suppresses tumor growth and inhibits clonogenic cell survival of a model prostatic adenocarcinoma, the Dunning R3327G rat tumor. Cancer Res 53:4855–4859

Lou H, Yeager M, Li H, Bosquet JG, Hayes RB, Orr N, Yu K, Hutchinson A, Jacobs KB, Kraft P, Wacholder S, Chatterjee N, Feigelson HS, Thun MJ, Diver WR, Albanes D, Virtamo J, Weinstein S, Ma J, Gaziano JM, Stampfer M, Schumacher FR, Giovannucci E, Cancel-Tassin G, Cussenot O, Valeri A, Andriole GL, Crawford ED, Anderson SK, Tucker M, Hoover RN, Fraumeni JF Jr, Thomas G, Hunter DJ, Dean M, Chanock SJ (2009) Fine mapping and functional analysis of a common variant in MSMB on chromosome 10q11.2 associated with prostate cancer susceptibility. Proc Natl Acad Sci USA 106(19):7933–7938

Manolio TA, Brooks LD, Collins FS (2008) A HapMap harvest of insights into the genetics of common disease. J Clin Invest 118:1590–1605

Mundle SD, Sheth NA (1993) Suppression of DNA synthesis and induction of apoptosis in rat prostate by human seminal plasma inhibin (HSPI). Cell Biol Int 17:587–594

Nam RK, Reeves JR, Toi A, Dulude H, Trachtenberg J, Emami M, Daigneault L, Panchal C, Sugar L, Jewett MA, Narod SA (2006) A novel serum marker, total prostate secretory protein of 94 amino acids, improves prostate cancer detection and helps identify high grade cancers at diagnosis. J Urol 175:1291–1297

Peng Y, Li CX, Chen F, Wang Z, Ligr M, Melamed J, Wei J, Gerald W, Pagano M, Garabedian MJ, Lee P (2008) Stimulation of prostate cancer cellular proliferation and invasion by the androgen receptor co-activator ARA70. Am J Pathol 172:225–235

Reeves JR, Dulude H, Panchal C, Daigneault L, Ramnani DM (2006) Prognostic value of prostate secretory protein of 94 amino acids and its binding protein after radical prostatectomy. Clin Cancer Res 12:6018–6022

Rothberg JM, Leamon JH (2008) The development and impact of 454 sequencing. Nat Biotechnol 26:1117–1124

Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol 132:365–386

Stanbrough M, Bubley GJ, Ross K, Golub TR, Rubin MA, Penning TM, Febbo PG, Balk SP (2006) Increased expression of genes converting adrenal androgens to testosterone in androgen-independent prostate cancer. Cancer Res 66:2815–2825

Sun J, Zheng SL, Wiklund F, Isaacs SD, Purcell LD, Gao Z, Hsu FC, Kim ST, Liu W, Zhu Y, Stattin P, Adami HO, Wiley KE, Dimitrov L, Li T, Turner AR, Adams TS, Adolfsson J, Johansson JE, Lowey J, Trock BJ, Partin AW, Walsh PC, Trent JM, Duggan D, Carpten J, Chang BL, Gronberg H, Isaacs WB, Xu J (2008) Evidence for two independent prostate cancer risk-associated loci in the HNF1B gene at 17q12. Nat Genet 40:1153–1155

The International HapMap Consortium (2005) A haplotype map of the human genome. Nature 437:1299–1320

Thierry-Mieg D, Thierry-Mieg J (2006) AceView: a comprehensive cDNA-supported gene and transcripts annotation. Genome Biol 7(Suppl 1):S12.1–S14

Thomas G, Jacobs KB, Yeager M, Kraft P, Wacholder S, Orr N, Yu K, Chatterjee N, Welch R, Hutchinson A, Crenshaw A, Cancel-Tassin G, Staats BJ, Wang Z, Gonzalez-Bosquet J, Fang J, Deng X, Berndt SI, Calle EE, Feigelson HS, Thun MJ, Rodriguez C, Albanes D, Virtamo J, Weinstein S, Schumacher FR, Giovannucci E, Willett WC, Cussenot O, Valeri A, Andriole GL, Crawford ED, Tucker M, Gerhard DS, Fraumeni JF Jr, Hoover R, Hayes RB, Hunter DJ, Chanock SJ (2008) Multiple loci identified in a genome-wide association study of prostate cancer. Nat Genet 40(3):310–315

Vanaja DK, Cheville JC, Iturria SJ, Young CY (2003) Transcriptional silencing of zinc finger protein 185 identified by expression profiling is associated with prostate cancer progression. Cancer Res 63:3877–3882

Yeager M, Orr N, Hayes RB, Jacobs KB, Kraft P, Wacholder S, Minichiello MJ, Fearnhead P, Yu K, Chatterjee N, Wang Z, Welch R, Staats BJ, Calle EE, Feigelson HS, Thun MJ, Rodriguez C, Albanes D, Virtamo J, Weinstein S, Schumacher FR, Giovannucci E, Willett WC, Cancel-Tassin G, Cussenot O, Valeri A, Andriole GL, Gelmann EP, Tucker M, Gerhard DS, Fraumeni JF Jr, Hoover R, Hunter DJ, Chanock SJ, Thomas G (2007) Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. Nat Genet 39:645–649

Yeager M, Xiao N, Hayes RB, Bouffard P, Desany B, Burdett L, Orr N, Matthews C, Qi L, Crenshaw A, Markovic Z, Fredrikson KM, Jacobs KB, Amundadottir L, Jarvie TP, Hunter DJ, Hoover R, Thomas G, Harkins TT, Chanock SJ (2008) Comprehensive resequence analysis of a 136 kb region of human chromosome 8q24 associated with prostate and colon cancers. Hum Genet 124:161–170

Yeh S, Chang C (1996) Cloning and characterization of a specific coactivator, ARA70, for the androgen receptor in human prostate cells. Proc Natl Acad Sci USA 93:5517–5521